Modeling and analysis of a bacterial stochastic switch

Brian Munsky and Mustafa Khammash

***Abstract*— Deterministic models fail to capture certain important dynamics in the subcellular environment due to the discrete stochastic nature of the molecular interactions at the gene level. Such discrete stochastic interactions are exploited in the cell to implement stochastic switches whose state are predictable only in a statistical sense. This absence of determinism and the inherent variability resulting from it play an important role in creating biological diversity that improves the chance for survivability. This paper will use a simplified model of the Pap switch in *E. coli* in order to illustrate a variety of computational methodologies. It will be shown that continuous time discrete-state Markov chains are natural tools for modeling this switch, and a review of these approaches will be provided. Using the recently introduced Finite State Projection algorithm, it is shown that the probability of a given switch position can be computed within any a-priori tolerance without resorting to the Monte-Carlo simulations, which generally lack accuracy guarantees.**

I. Introduction

Modern genetic and molecular biology techniques have successfully revealed elaborate regulatory networks, which ultimately control various biologic responses. A major goal of Systems Biology is to utilize this wealth of information to produce predictive models of how individual regulatory steps integrate to produce observable responses. Such models could assist biologists and biochemists to (1) understand complicated regulatory phenomena, (2) pinpoint key regulatory features and alter those features to achieve desired outcomes, and (3) achieve better understanding of how and why regulatory systems have evolved in different species.

One of the major difficulties encountered in achieving good predictive models is that cellular processes often involve random molecular events and are subject to vast amounts of signal noise. In spite of this randomness, the events and processes that make up cellular function are highly ordered and often tightly regulated. Random fluctuations within the cell are frequently suppressed through cellular networks that employ feedback and implement cellular "filters." These low pass filters suppress fast variations but admit the slowly varying components of cellular signals. In other important cases, however, randomness works toward the cell's advantage. Isogenic heterogeneity directly results from a random step that is exploited specifically to that end; for example, the lysis/lysogeny decision in bacteriophage lambda (phage), utilizes randomness to operate a genetic switch [1]. Here, a homogeneous and genetically identical phage population takes two separate paths with one fraction of the population following a lytic fate, while the other following a lysogenic fate.

Understanding the impact of random fluctuations in large cellular networks, presents one of the most challenging questions in systems biology. Deterministic models of biochemical networks have focused on tracking the concentrations of key reactants (e.g. proteins, mRNA molecules, etc.) within the cell as a function of time. Here reaction rate equations are derived based on Michaelis-Menten type kinetics that result in a system of nonlinear differential equations. When reactant species are present in large numbers, fluctuations may not be important, and their mean behavior can be accurately captured by their concentration within the cell. However, many biological networks frequently involve molecules that are present in small numbers. In these cases random fluctuations may be amplified significantly and have great impact on the cell. Deterministic models are generally not adequate for modeling such fluctuations. Furthermore, continuous representations of very small reactant quantities fail to capture the discrete behavior of such reactants. Instead, these processes must be modeled using discrete and stochastic models.

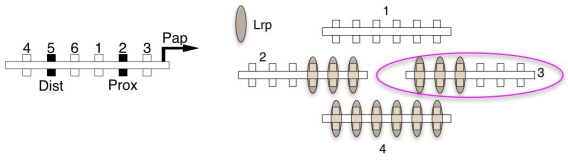
As stochastic modeling of biological processes becomes more important to the scientific community, more powerful computational tools are needed. Many Monte Carlo methods have been proposed to meet these needs including the Stochastic Simulation Algorithm[2], [3], various τ leaping approaches [4], [5], [6], [7], and systems partitioning methods [8], [9], [10], [11]. More recently, the authors have proposed the Finite State Projections algorithm [12], which directly approximates to the evolution of the probability distributions without computing individual realizations. In this paper we will evaluate a few of these computational methods on a simplified stochastic model of the pyelonephritis-associated pili (Pap) system controlling expression of Pap pili in *E. coli*.

The next section provides a brief description of the Pap model. In the section 3 we explain how a few numerical solution schemes may be applied to the Pap switch. In section 4 we compare and contrast the results found using each of the schemes. Finally, we provide a few concluding remarks regarding the efficiency and accuracy of the different computational methods in regard to the Pap switch model.

II. Modeling

The model of the Pap system to be considered in this paper is relatively simple with a single operon and three regulatory factors: leucine-responsive regulatory protein (Lrp), DNA adenine methylase (Dam), and the local *pap*-encoded regulatory protein PapI.

The *pap* operon(Fig. 1a) provides the basic structure of the switch and defines the rules of all regulatory actions. The operon contains six *pap* DNA Lrp binding sites spaced three helical turns apart We have designated the sites 1 to 6, where



(a) (b)

Fig. 1. (a) The regulatory region of the *pap* operon. (b) Schematic of the four different possible Lrp binding patterns.

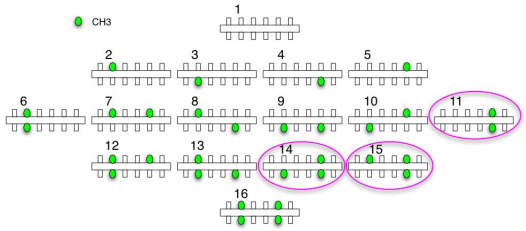


Fig. 2. Schematic of the 16 different possible methylation patterns.

1-2-3 correspond to the sites proximal to the *papB* promoter, and sites 4-5-6 are those distal to the *papB* promoter. The regulatory region also contains four DNA sites with the genetic sequence GATC. Two of these occur at Lrp binding site 2 (top and bottom strand), designated $GATC_{prox}$. The other two occur at site 5, designated $GATC_{dist}$. These sites are targets for Dam, which places a methyl group on the adenine of the GATC sequence.

Lrp binds cooperatively and reversibly in sets of three dimers at the proximal or distal sites forming four possible Lrp binding patterns (see Fig. 1b). The affinity of each pattern depends upon the methylation pattern of the operon and on the population of PapI in the system[13]. Mutational analyses shows that disrupting Lrp binding site 2 increases *pap* activation, and disrupting Lrp at site 5 decreases *pap* activity[14]. Thus, we assume that *pap* is transcribed only when:

(1) Lrp has bound to the distal sites, and

(2) Lrp has not bound to the proximal sites (see Fig. 1b, circled state).

Dam irreversibly methylates the *pap* operon in four GATC locations: top-prox, bottom-prox, top-dist and bottom-dist. As a result, the operon can achieve sixteen different methylation patterns (See Fig. 2). Examination of the pap DNA methylation patterns showed that in phase ON cells $GATC_{prox}$ is methylated and $GATC_{dist}$ is nonmethylated [15]. Based upon this information, we add the additional requirements for *pap* transcription:

(3) Dam has methylated the proximal sites (top and bottom), and

(4) Dam has not fully methylated the distal site (see Fig. 2, circled states).

Combining the Lrp binding patterns with the Dam methylation patterns, there are 64 mutually exclusive *pap* operon configurations. Let the vector $\mathbf{g} = [g_1, g_2, \ldots, g_{64}]^T$ denote the populations of all configurations, $\mathbf{g}$ is a vector of zeros except for a single value of one. The operon changes from the i^{th} configuration to another through 192 different stochastic reactions. 64 of these reactions, $\mu = \{1, 2, \ldots, 64\}$, are bimolecular[1] Lrp association events where Lrp binds to the operon at sites 1-2-3 or 4-5-6. The propensity function, a_μ, is a measure of the liklihood of these reactions and depends on the current configuration and the population of Lrp: $a_\mu = c_\mu[g_i][Lrp]$, where c_μ is the reaction rate and $[g_i]$ and $[Lrp]$ denote the integer populations of the i^{th} operon configuration and Lrp, respectively. There are 64 monomolecular[2] reaction events, $\mu = \{65, 65, \ldots, 128\}$, in which Lrp dissociates from the operon; these have propensity functions given by: $a_\mu = c_\mu[g_i]$. Finally there are 64 bimolecular Dam methylation events, $\mu = \{129, 130, \ldots, 192\}$ with propensity function given by: $a_\mu = c_\mu[g_i][Dam]$. Of the 64 operon configurations, only three satisfy requirements 1-4 above and can result in Pap transcription. When in one of these three configurations, the local protein PapI is produced in a Poisson pure birth process with constant propensity function $a_b = c_b$. PapI can also degrade in a monomolecular event with propensity function: $a_d = c_d[r]$, where $[r]$ denoted the population of PapI. In turn, PapI increases the affinity of Lrp for the distal site and creates a positive feedback loop: $c_\mu = c_\mu(r)$ for $\mu \in \{65, 66, \ldots, 128\}$.

For a known amount of Dam and Lrp in the system, the vector pair $\mathbf{x}_i = (g_j, r)$, defines the state of the system. We can enumerate all of the possible states as $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ In this framework, chemical reactions are simply jumps from $\mathbf{x}_i$ to $\mathbf{x}_j$. The direction of the μ^{th} jump is given by the reaction stoichiometry, $\mathbf{x}_j = \mathbf{x}_i + \nu_\mu$. It has been shown that if a chemically reacting system is well-mixed and has a fixed volume and fixed temperature, then that system is a Markov process[16], [17], and the waiting times between each reaction are exponentially distributed with mean equal to the inverse of the propensity function give by $a_\mu = a_\mu(\mathbf{x}_i)$. For our model, we assume that the system begins in the state $\mathbf{x}_1 = [g_1, 0]$: there is no prior methylation, there is no Lrp bound to the operon, and there is no PapI in the system. For the purposes of this model, we define the system to be ON if there are more than 20 molecules of PapI in the system; otherwise the system is considered to be OFF. For our enumeration, this definition corresponds to the system being in any state $\mathbf{x}_i$ where $i > 1280$. We are interested in determining the probability that the system is ON as a function of time. In the next section we discuss how one would obtain these results for various computational methodologies.

III. Numerical Methods

Under our assumptions, the Pap system behaves as a continuous time, discrete space Markov process with exponentially distributed waiting times. We are interested in computing the probabilities of the various possible system

[1]Involve two reactive species

[2]Involve only one reactive species

states as functions of time. For the states $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$, we define the corresponding probability density state vector as $\mathbf{P} := [P_1, P_2, \ldots]^T$. The evolution of $\mathbf{P}$ over time can be described by the possibly infinite dimensional set of linear, time invariant ordinary differential equations known as the Chemical Master Equation (CME) [16], [17]: $\dot{\mathbf{P}} = \mathbf{AP}$. In most cases the CME has not been directly solved, and analyses are often conducted using Monte Carlo algorithms. In the next subsection, we will discuss a few of these algorithms in relation to the Pap switch model.

A. Monte Carlo Algorithms

The first and perhaps most widely used Monte Carlo algorithm for stochastic chemical kinetics is Gillespie's Stochastic Simulation Algorithm (SSA)[2], [3]. In the SSA, one simulates the chemical process one reaction event at a time. The basic idea is as follows. At any time, t, the system is at some specific state, $\mathbf{x}(t) = \mathbf{x}_i$. There are at most M possible reaction that the system can make beginning at $\mathbf{x}_i$, each with a propensity depending solely on $\mathbf{x}_i$: $a_\mu = a_\mu(\mathbf{x}_i)$ for $\mu = \{1, 2, \ldots, M\}$. The time until the next reaction, τ, is generated from a single exponentially distributed random variable with a mean equal to $\left(\sum_{\mu=1}^{M} a_\mu(\mathbf{x}_i)\right)^{-1}$. Once τ has been generated, one can generate a second random number and the propensity functions to choose which of the M reactions occurs. The time of the simulation is incremented to $t + \tau$ and the state of the system is updated by the stoichiometry of the chosen reaction. The process is continued until the final time of interest.

Although the SSA produces detailed realizations for stochastically evolving chemical systems, the method becomes very computationally expensive when the system undergoes enormous numbers of individual reactions. In these cases it is often necessary to sacrifice some of the precision of the SSA for faster, yet approximate methods such as time-leaping methods and system-partitioning methods.

B. τ leaping methods

Time-leaping methods rely on the τ *leap assumption* that many reaction events will occur in a period of time without significantly changing the reaction propensity functions [18]. Using this assumption, one can simulate the reactions during each time step as a set of independent pure birth (Poisson) processes. A major difficulty for the τ-leaping method results when too many critical reactions are included in a single leap such that some molecular populations become negative. More recent versions of τ-leaping, including binomial τ leaping [7] and adaptive and implicit τ leaping [4], [5], [6] are more robust than their predecessors, but their accuracy remains severely compromised when very small populations of interacting molecules result in fast, dramatic changes in propensity functions. For the Pap model, this is a particularly important concern. The majority of the reactions involve the *pap* operon, which is present in the cell in a single copy, and nearly every reaction violates the τ leap assumption. As a result τ leaping cannot be used for the Pap model.

C. System partitioning methods

The second approach to speeding up the SSA involves separating the system into slow and fast partitions, approximating the dynamics of the fast partition, and then stochastically simulates the remaining slow partition.[8], [9], [10], [11] For such a scheme to work, it is necessary that there is a sufficient gap between the between fast and slow reactions. Depending upon the parameters involved in the Lrp binding and unbinding reactions, the Pap model described above may be amenable to such an approach. For example, if the Lrp reactions were sufficiently fast, we could assume that the four Lrp binding configurations (shown in Figure 2) would reach a probabilistic steady state before any other event (Dam methylation, PapI production, or PapI degradation) may occur. Thus, under this assumption we could algebraically compute a distribution for the four Lrp binding configurations as a function of the 16 Dam methylation patterns and PapI population. By skipping the individual Lrp events and concentrating on the other stochastic events, we significantly reduce the number of events in each simulation.

For any of these Monte Carlo methods, a statistical description of the system's dynamics, such as the probability density, mean, or variance, requires a large number of realizations. To analyse rare, but important biological traits, extremely large numbers of simulations may be required. For example, the Pap Pili epigenetic switch in *E. coli* has an OFF to ON switch rate on the order of 10^{-4} per cell per generation [19]. In order to capture this switch rate with a relative accuracy of one percent, one would require far more than one million Monte Carlo simulations.

D. The Finite State Projection Method

Recently, the authors proposed a promising new algorithmic approach to solving the CME: the Finite State Projection (FSP) algorithm[12]. This algorithm provides an analytical approximation to the CME without generating Monte Carlo simulations and includes a guarantee as to its own accuracy.

As stated above the probability distribution of the system evolves according to a set of linear ordinary differential equations: $\dot{\mathbf{P}}(t) = \mathbf{AP}(t)$, where $\mathbf{P}(t)$ is a vector of the probabilities $[P_1(t), P_2(t), \ldots]$ of every possible state $[\mathbf{x}_1, \mathbf{x}_1, \ldots]$, and $\mathbf{A}$ is the generator matrix that expresses the stochastic rates of transition from one state to another. The columns and rows of $\mathbf{A}$ are uniquely defined by the reaction stoichiometries and propensities and the choice of the enumeration of our state space. $\mathbf{A}$ has some very useful properties: it is independent of t; all of its diagonal elements are non-positive; all its off-diagonal elements are non-negative; and all its columns sum to zero.

If the system of equations were finite, then the solution could be found using a matrix exponential. However, $\mathbf{P}(t)$ has infinite dimension–the problem cannot generally be solved exactly, but a projection may be made to achieve an arbitrarily accurate approximate solution. It is helpful to introduce some convenient notation. Let $J = \{j_1, j_2, j_3, \ldots\}$ denote an index set. For any vector $\mathbf{v}$ let $\mathbf{v}_J$ denote the subvector of $\mathbf{v}$ whose elements are chosen according to J,

and for any matrix $\mathbf{A}$, let $\mathbf{A}_J$ denote the submatrix of $\mathbf{A}$ such that the rows and columns have been chosen according to J. For example, if J is defined as $\{3,1\}$, then:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}_J = \begin{bmatrix} k & g \\ c & a \end{bmatrix}.$$

In a previous work, we stated the following theorem[3].
Theorem 2.1 Consider any Markov process in which the probability density state vector evolves according to the linear ODE:

$$\dot{\mathbf{P}}(t) = \mathbf{A} \cdot \mathbf{P}(t),$$

If for some finite index set J, $\varepsilon > 0$, and $t_f \geq 0$,

$$\mathbf{1}^T \exp(\mathbf{A}_J t_f)\mathbf{P}_J(0) \geq 1 - \varepsilon, \quad (1)$$

then

$$\exp(\mathbf{A}_J t_f)\mathbf{P}_J(0) \leq \mathbf{P}_J(t_f) \leq \exp(\mathbf{A}_J t_f)\mathbf{P}_J(0) + \varepsilon\mathbf{1}. \quad (2)$$

Using this FSP theorem, if we can find a principle submatrix $\mathbf{A}_J$ of $\mathbf{A}$, such that:

$$\mathbf{1}^T \exp(\mathbf{A}_J t)\mathbf{P}_J(0) \geq 1 - \varepsilon,$$

then we are guaranteed that the probability density for every state at time, $t = t_f$, is within the bounds given by:

$$\exp(\mathbf{A}_J t)\mathbf{P}_J(0) \leq \mathbf{P}_J(t) \leq \exp(\mathbf{A}_J t)\mathbf{P}_J(0) + \varepsilon\mathbf{1}.$$

We can use an algorithmic approach to add more and more states to the finite projection until we obtain an error, ε, that is less than a prespecified bound.

Like the SSA, the FSP algorithm is amenable to some time-partitioning approximation schemes. For example, if the Lrp binding/unbinding reactions are very fast in the Pap model, one can assume that these four configurations instantaneously reach probabilistic equilibrium and thereby reduce the order of the CME by a factor of four.

In the next section, we solve the CME for the Pap problem with two different computational methods: the SSA and the FSP. For each method we show results with and without making the assumption that Lrp binding/unbinding reaches probabilistic equilibrium.

IV. Results and Discussion

Fig. 3 illustrates the probability that the system is ON (contains at least 20 molecules of PapI) as a function of time for three different sets of Lrp binding/unbinding rates: slow, moderate, and fast. The three solid curves have been obtained using the FSP formulation with an error tolerance of 10^{-5}. A fourth dashed curve illustrates the results using the FSP with the assumption that Lrp binding/unbinding reaches probabilistic equilibrium before any other reaction. The figure shows that as the Lrp reactions become faster, the dynamics of the system approaches the approximate solution, suggesting that there is some threshold in Lrp binding/unbinding rates above which the assumption of probabilistic equilibrium is valid. Below that threshold, however, the equilibrium assumption is not valid and the full SSA or full FSP methods must be used.

[3]For proof see Munsky and Khammash, 2006

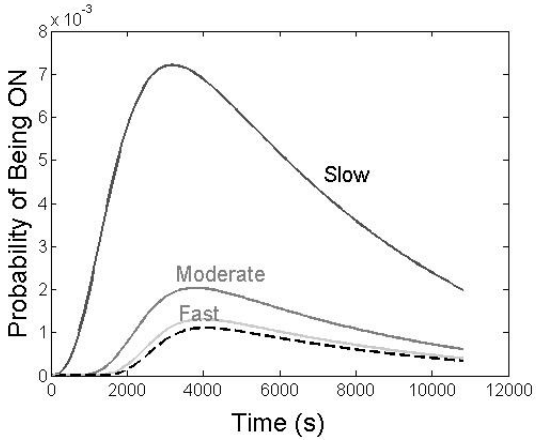


Fig. 3. Proportion of ON cells versus time for three sets of Lrp reaction rates. Solid lines are computed using the FSP with an error tolerance of 10^{-5}; the dashed line is computed using the FSP with an assumption of probabilistic equilibrium on the Lrp association/dissociation reactions.

TABLE I
A comparison of the efficiency and accuracy of the FSP, SSA, and approximate FSP and SSA methods.

Method	# Simulations	Time (s)[a]	Relative Error[b]
FSP	Does not apply. [c]	42.1	$< 0.013\%$
FSP approx.	Does not apply.	3.3	$\approx 1.3\%$
SSA	10^4	$>$ 26 days	Not available
SSA approx.	10^4	9.8	$\approx 16\%$
SSA approx.	10^5	94.9	$\approx 7.7\%$
SSA approx.	10^6	946.2	$\approx 1.6\%$

[a]All computations have been performed in Matlab 7.2 on a 2.0 MHz PowerPC G5.
[b]Error in switch rate is computed at $t = 4000s$
[c]The FSP is run only once with a specified allowable total error of 10^{-5}.

Fig. 4 shows the results for the fast Lrp reaction rates as computed with the full FSP (dashed line), and the approximations of both the SSA (three gray lines corresponding to sets of 10^4, 10^5 and 10^6 simulations) and the FSP (solid black line)[4]. From the figure one can observe that as the number of runs increases, the prediction of the approximate SSA approaches that of the approximate FSP solution; however, even after a million runs the SSA solution, the error in the SSA solution is yet visible. Table I provides relative errors errors in the probability of being ON at time $t = 4000$ for the FSP, the approximate FSP and the approximate SSA algorithms. For reference the computational time of each method is also included. In terms of accuracy, the FSP vastly outperforms the other two algorithms.

[4]For the fast Lrp reaction rates, a single realization of the full SSA requires more than 230 seconds. 10^5 realizations requires a total simulation time of more than 260 days.

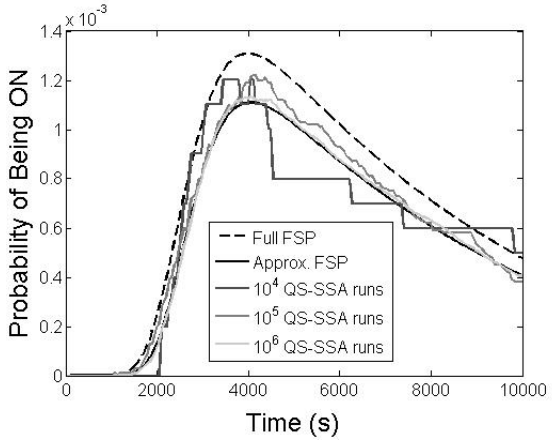


Fig. 4. Proportion of ON cells versus time for fast Lrp reaction rates. The dashed line corresponds to the FSP computation with an error tolerance of 10^{-5}. The solid lines are computed with an assumption of probabilistic equilibrium on the Lrp reactions. The smooth black line is computed using the approximate FSP algorithm, and the gray lines correspond to computations with the approximate SSA. Se also Table I.

V. Conclusions

This paper has outlined a simplified model of the Pap switch in *E. coli*, and has reviewed a few numerical procedures for the numerical solution of the Chemical Master Equation that arises in this model. Because the Pap switch occurs on the order of once per thousand cells per generation, great computational precision is necessary. Two types of numerical solutions were considered: Monte Carlo algorithms and the Finite State Projection algorithm. To obtain meaningful statistical results, Monte Carlo algorithms require the generation of a vast number of individual realizations. For the Pap model, a single run of the Stochastic Simulation Algorithm (SSA) takes nearly five minutes, and the necessary precision is unobtainable in a reasonable amount of time. Furthermore, because the majority of the reactions involve a gene that is present as a single copy within the cell, every individual reaction is important, and τ leaping algorithms offer no advantage over the basic SSA. Time partitioning approximations of the SSA, in which some reactions are assumed to reach instantaneous probabilistic equilibrium, can significantly speed up the computation, but are only valid for some sets of parameters.

As an attractive alternative to Monte Carlo algorithms, the FSP directly computes the system's probability density vector at a given time without requiring the computation of large numbers of process realizations. The FSP automatically guarantees upper and lower bounds on the solution of the true system. For some sets of parameters in the Pap model, where the SSA takes too long, τ leaping is invalid, and system partitioning methods provide inaccurate results, the FSP may well be the only feasible tool. For other sets of parameters, where system partitioning methods are applicable, the FSP can benefit just as well as does the SSA.

References

[1] A. Arkin, J. Ross, and McAdams H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics*, 149:1633–1648, 1998.

[2] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2360, May 1977.

[3] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 104:1876–1889, 2000.

[4] D. T. Gillespie and L. R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.*, 119(16):8229–8234, Oct. 2003.

[5] Y. Cao, D. T. Gillespie, and L. R. Petzold. Avoiding negative populations in explicit poisson tau-leaping. *J. Chem. Phys.*, 123(054104), 2005.

[6] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *J. Chem. Phys.*, 119(24):12784–12794, Dec. 2003.

[7] T. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *J. Chem. Phys.*, 121(21):10356–10364, Dec. 2004.

[8] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *J. Chem. Phys.*, 118(11):4999–5010, Mar. 2003.

[9] E. Haseltine and J. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117(15):6959–6969, Jul. 2002.

[10] Y. Cao, D. Gillespie, and L. Petzold. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.*, 122(014116), Jan. 2005.

[11] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biological reactions. *J. Chem. Phys.*, 112(054103), 2005.

[12] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(044104), 2006.

[13] A. D. Hernday, B. A. Braaten, and D. A. Low. The mechanism by which dna adenine methylase and papi activate the pap epigenetic switch. *Mol. Cell*, 12:947–957, October 2003.

[14] X. Nou, B. A. Braaten, L. Kaltenbach, and D. A. Low. Differential binding of lrp to two sets of *pap* dna binding sites mediated by pap i regulates pap phase variation in *escherichia coli*. *EMBO Journal*, 14:5785–5797, 1995.

[15] B. A. Braaten, X. Nou, L. S. Kaltenbach, and D. A. Low. Methylation patterns in *pap* regulatory dna control pyelonephritis-associated pili phase variation. *Cell*, 76:577–588, 1994.

[16] D. McQuarrie. Stochastic approach to chemical kenetics. *J. Applied Probability*, 4:413–478, 1967.

[17] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.

[18] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115(4):1716–1733, Jul. 2001.

[19] L. B. Blyn, B. A. Braaten, C. A. White-Ziegler, D. H. Rolfson, and D. A. Low. Phase-variation of pyelonephritis-associated pili in escherichia coli: Evidence for transcriptional regulation. *EMBO J.*, 8:613–620, 1989.